

# A novel data pre-processing method for odour detection and identification system

Wentian Zhang<sup>a</sup>, Taoping Liu<sup>a</sup>, Lin Ye<sup>a</sup>, Maiken Ueland<sup>b</sup>, Shari L. Forbes<sup>b</sup>, Steven W. Su<sup>a,\*</sup>

<sup>a</sup>*Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia.*

<sup>b</sup>*School of Mathematical and Physical Sciences, University of Technology Sydney, Sydney, NSW 2007, Australia*

---

## Abstract

This paper presents a novel electronic nose (E-nose) data pre-processing method, based on a recently developed non-parametric kernel-based modelling (KBM) approach. The proposed method is tested by an automated odour detection and classification system, named "NOS.E", developed by the NOS.E team in University of Technology Sydney. Experimental results show that when extracting the derivative-related features from signals collected by the NOS.E, the proposed non-parametric KBM odour data pre-processing method achieves more reliable and stable pre-processing results comparing with other pre-processing methods such as wavelet package correlation filter (WPCF), mean filter (MF), polynomial curve fitting (PCF) and locally weighted regression (LWR). Based on these derivative-related features, the NOS.E can achieve a 96.23% accuracy of classification with the popular Support Vector Machine (SVM) classifier.

**Keywords:** Electronic nose (E-nose), Instrumentation, Data pre-processing, Non-parametric kernel-based modelling method

---

## 1. Introduction

Electronic nose (E-nose) is capable of identifying chemical compounds through sensing and analysing odour molecules. As a kind of machine olfaction, E-nose plays a significant role in the odour analysis area and has received considerable attention from researchers all over the world [1–4]. The E-nose system comprises a set of active gas sensors that detect the odour and transduce the chemical vapours into electrical signals [1, 2]. The odour "fingerprint" captured by the gas sensors can then be analysed and identified with pattern classification methods, e.g., Principal Components Analysis (PCA), Cluster Analysis (CA), Support Vector Machine (SVM), and Artificial neural networks (ANNs). E-nose has been extensively applied in the areas of agriculture, medical diagnosis, environmental monitoring and protection, food safety, the military, cosmetics and pharmaceuticals [1, 4–11].

Currently, by using different E-nose platforms, the studies on E-nose mainly focus on two different parts: 1. The design of hardware system (such as sensor design and main control system design) [9, 11–17]; 2. The algorithms for E-nose, such as data pre-processing methods and odour classification methods [9, 18–25]. Moreover, some researchers develop their E-nose research based on the famous commercial E-nose products (such as the *fox* E-nose (*Alpha MOS, France*), the portable *Cyranose 320* (*Cyranose Science, USA*), *Airsense PEN2* and *PEN3* (*Airsense Analytics GmbH, Germany*) [1, 26–32]. These studies on E-nose have made great progress in this area.

In order to get reliable classification results for E-nose applications, data pre-processing methods are used to improve the stability of the feature extracted from the pre-processed odour data. These pre-processing techniques mainly include wavelet package correlation filter (WPCF), mean filter (MF),

polynomial curve fitting (PCF) and locally weighted regression (LWR), etc. [33, 34]. Even though these methods are quite mature and efficient, sometimes they are unable to obtain reliable results due to individual variations in the test system and unexpected responses caused by the gas interference or fluctuations of environmental parameters [35–37]. Moreover, these unexpected responses treated as noises will potentially reduce the stability and reliability of features. Experimental results will be significantly influenced especially for the derivative-related features which are sensitive to noises.

To seek a data pre-processing method which can overcome the drawbacks of current data pre-processing techniques for E-nose system, this paper proposes a novel non-parametric kernel-based modelling (KBM) data pre-processing method. Furthermore, this method is tested by recently developed NOS.E odour detection and identification system. The NOS.E system (shown in Fig. 1), mainly comprises an efficient power system, an automated air intake system, an interchangeable metal-oxide (MOX) gas sensor array board, and a fast data acquisition module. The target odour is drawn into the mixing chamber by the gas sampling pump, before going into the gas chamber, where the sensor array senses the odour stimulus.

Before applying the non-parametric KBM approach [38] [39] in the area of E-nose data pre-processing, three standard signals (linear signal, logarithmic signal and sigmoid signal) are used to test the performance of five different pre-processing methods (WPCF, MF, PCF and LWR). The results (presented in Section 3) show that the proposed method provides more reliable pre-processing results comparing with other methods. Moreover, these data pre-processing methods are applied on real odour signals collected by the NOS.E system and then the Coefficient of Variation (CoV) method is employed to evaluate the stability of derivative-related features [40]. The CoV analysis results (listed in Table 2) for different perfumes features indicate that the proposed method is much better than other data pre-processing methods.

---

\*Corresponding author.

Email address: [Steven.Su@uts.edu.au](mailto:Steven.Su@uts.edu.au) (Steven W. Su)

In addition, when applying the popular SVM classifier to the obtained features, 96.23% accuracy of classification is achieved based on the proposed method. On the other hand, the WPCF data pre-processing method has 94.12% accuracy of classification which is the second highest compared with other data pre-processing methods (68.57%, 87.80%, and 71.50% accuracy of classification for MF, PCF, and LWR data pre-processing methods, respectively).

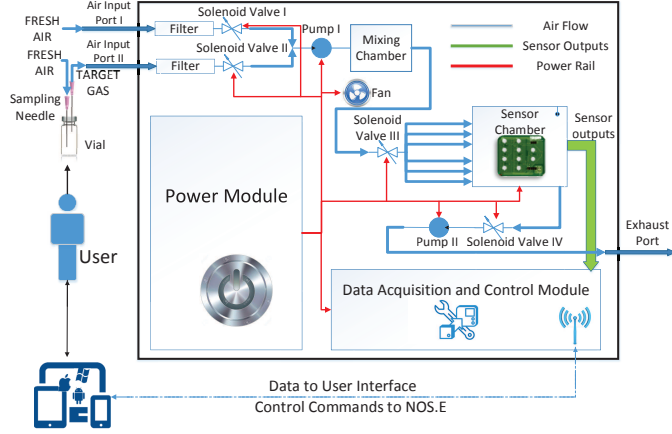


Figure 1: Block Diagram of NOS.E System.

The paper is structured as follows. Section 2 is devoted to the experimental setup. Data pre-processing methods are developed in section 3. Section 4 provides experimental results and analysis discussion. The conclusions are drawn in section 5.

## 2. Experimental Setup

### 2.1. Data Pre-processing Simulation Setup

In this paper, three different standard simulated test signals (as shown in Fig 2): linear signal (Lin)  $y = x$ ; logarithmic signal (Log)  $y = \log(x)$  and sigmoid signal (sig)  $y = \frac{1}{(1+\exp(x))}$  are used to test the performance of five different pre-processing methods (KBM, WPCF, MF, LWR and PCF). Firstly, these signals are polluted by Gaussian noises with the Signal-to-Noise Ratio (SNR) ranging from 10dB to 50dB. Then the polluted signals are processed by the five different data pre-processing methods. Finally, the Normalized Root Mean Square Error (NRMSE) is used as a criterion to determine the goodness-of-fit between the pre-processed signals and the original signal, where the NRMSE costs vary between -Inf (bad fit) to 1 (perfect fit).

### 2.2. Perfume Test Experimental Setup

In this paper, the NOS.E system are used to detect two different kinds of perfume samples (bought from David Jones, Bondi Junction, Sydney, Australia) to verify the performance of the proposed data pre-processing method. The diagram of NOS.E perfume test experimental setup is shown in Fig. 3. Users can operate the NOS.E system via a touchscreen Pad. The integral components of this system are assembled in a carrying case. The sensor array in NOS.E equipment is composed of ten commercially available metal oxide gas sensors: *TGS 2611E*, *TGS 2612*, *TGS 2610D*, *TGS 2611C*, *TGS 2610C*, *TGS 2602*, *TGS 2600*, *TGS 2620*, *TGS 2603*, and *TGS 2602*. The

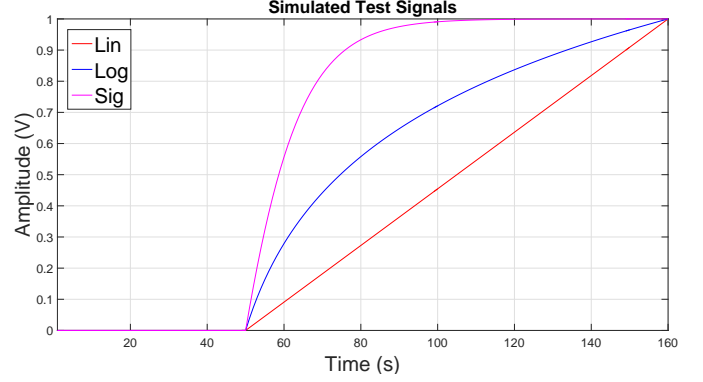


Figure 2: Simulated Test Signals.

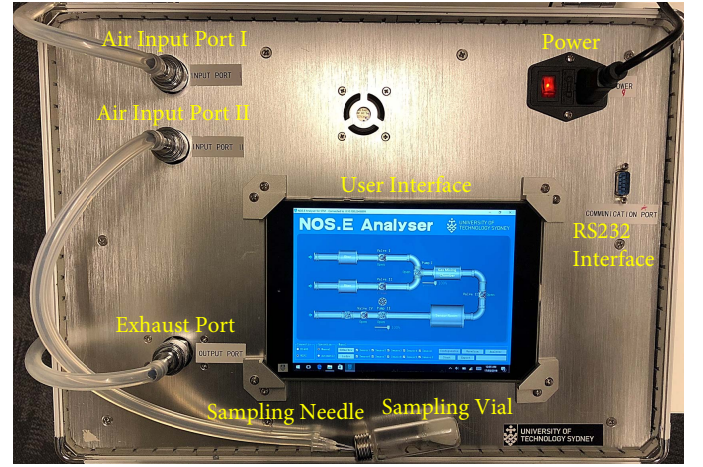


Figure 3: Diagram of NOS.E Perfume Test Experimental Setup.

input gases are sensed by the sensor array, converted to digital signals and then sent to the host computer for data processing.

In this paper, perfume I is *CHANEL Chance*, and perfume II is *CHANEL Gabrielle*. The test protocol of the NOS.E system are listed as following steps (using perfume I as an example).

1. Prepare the test sample;  
Prepare 1mL perfume using a 10mL headspace vial.
2. Power on NOS.E equipment and open all the actuators to warm up the NOS.E equipment for 60 minutes (to make sure the equipment works at the ideal working status);
3. Set up the test time for each phase using the NOS.E Analyzer;  
Chamber washing time for 300 seconds; Vacuum I for 10 seconds; Baseline setup for 20 seconds; Vacuum II for 10 seconds; Test time for 90 seconds; Baseline recovery for 90 seconds and Chamber washing II for 300 seconds.
4. Connect the test sample I;  
As shown in Fig. 1, the headspace vial has two sampling needles, one for the fresh air input and the other for the target gas output to NOS.E equipment.
5. Configuration;  
Click Configuration button on the user interface to send all the operating parameters and instructions to the slave computer.

- 6. Baseline set up;  
Click Baseline Set up button, and the NOS.E system will start to collect the data until the baseline recovery phase is complete.
- 7. Save the data;  
Once the test is completed, the data will be saved in the local folder automatically.
- 8. Review the test results;  
Click Analyser button to review the sensor responses and key features.
- 9. Repeat the same sample followed from steps 5 to 8 for ten times, and these repetitions are counted as ten acquisitions.

Followed by the NOS.E odour test protocol, 60 perfume samples (30 perfume I samples, 30 perfume II samples) are collected by the NOS.E system. Each sample is collected 10 times, hence, 600 odour datasets (300 perfume I datasets, 300 perfume II datasets) are analysed. The odour datasets used in this paper are obtained under 25 °C to 27 °C ambient temperature and 50% RH ambient humidity.

### 3. Methodology

In this section, a new non-parametric KBM method is exploited to model the gas sensor response [38, 39, 41]. Unlike most of the previous data pre-processing methods, in which the denoising is based on the filtering techniques, our approach reduces the noise based on non-parametric modelling to improve the classification accuracy; nevertheless, the identification of the model is based on the raw data.

Generally, the researchers use filtering based methods to process the raw data of the MOX gas sensor response. However, sometimes, the filter based pre-processing methods cannot obtain the desired effects due to the individual variation of the test system and the unexpected responses caused by the interference gas [35–37]. Especially, as the filter based methods cannot always guarantee the smoothness of the filtered signals, in feature extraction stage, unacceptable outliers might be generated. In order to obtain a better result and facilitate the automated feature extraction, we adopted a non-parametric modelling method which applied the finite impulse response to describe the systems characteristics.

In this paper,  $t$  with sampling time  $T$  is selected as the time index. The relationship between the gas input ( $u$ ), which can be approximately treated as a step stimulation, and the gas sensor response ( $y$ ) can be described by a single input single output (SISO) dynamic system. Hence, the discrete time output  $y$  can be calculated by the impulse response (IR) of this system by the following equation:

$$y(t) = \sum_{k=1}^{\infty} g_k^0 q^{-k} u(t) + \epsilon(t), \quad k = 1, 2, 3 \dots, \infty, \quad t = 1, 2, 3 \dots, N \quad (1)$$

where  $g_k^0$  represents the coefficient of the impulse response.  $q$  represents the shift operator, i.e.  $qu(t) = u(t+1)$ ,  $\epsilon(t)$  is the Gaussian white noise.

Considering the impulse response decays exponentially for linear stable systems, we here express the system by using the

$m$ th order finite impulse response (FIR) as:

$$G(q, \mathbf{c}) = \sum_{k=1}^m g_k q^{-k}, \quad \mathbf{c} = [g_1, g_2, \dots, g_m]^T. \quad (2)$$

Hence, the model in Eq.(1) is able to be transferred as:

$$y(t) = \boldsymbol{\varphi}^T(t) \mathbf{c} + \epsilon(t), \quad (3)$$

where  $\boldsymbol{\varphi}(t)$  contains the input information of the system:

$$\boldsymbol{\varphi}(t) = [u(t-1), u(t-2), \dots, u(t-m)]^T. \quad (4)$$

Then, the FIR model can be written as:

$$\mathbf{Y}_N = \boldsymbol{\phi}_N \mathbf{c} + \boldsymbol{\epsilon}_N, \quad (5)$$

where  $N = M - m$ , and  $M$  is the number of data point that we collected.  $\mathbf{Y}_N$  is a vector representation of sensor array's responses:

$$\mathbf{Y}_N = [y(1), y(2), \dots, y(t), \dots, y(N)]^T, \quad (6)$$

where  $y(t)$  denotes the  $t$ -th element of  $\mathbf{Y}_N$ .

$\boldsymbol{\epsilon}_N$  is a vector representation of Gaussian white noise:

$$\boldsymbol{\epsilon}_N = [\epsilon(1), \epsilon(2), \dots, \epsilon(t), \dots, \epsilon(N)]^T, \quad (7)$$

where  $\epsilon(t)$  denotes the  $t$ -th element of  $\boldsymbol{\epsilon}_N$ .

The  $i$ -th row of  $\boldsymbol{\phi}_N \in \mathbb{R}^{N \times m}$  is  $[u(m+i-1), u(m+i-2), \dots, u(i)]$ .

Assuming that function  $g \in \mathbb{R}^m$ , then function  $g$  in the regularisation term can be projected into a reproducing kernel Hilbert space (RKHS).

The IR model can be identified by minimising the cost function:

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^m} \|\mathbf{Y}_N - \boldsymbol{\phi}_N \mathbf{c}\|_2^2 + \gamma \mathbf{c}^T \mathbf{P}^{-1} \mathbf{c}, \quad (8)$$

where  $\mathbf{P}$  represents the kernel matrix which is defined as:

$$p(i, j) = e^{-\rho \|i-j\|^2}, \quad \rho > 0. \quad (9)$$

The estimated IR model from Eq.(8) can provide better and smoother results comparing to LASSO (Least Absolute Shrinkage and Selection Operator) or Ridge regression using the prior information in kernel matrix  $\mathbf{P}^{-1}$  [39]. Furthermore, as the system is stable, after a while, the impulse response will close to zero. Hence, when  $m$  is too big, we expect that the last several parameters of the estimated FIR approach to zero. Therefore, an extra  $\mathcal{L}_1$  regularisation was added to sparsify the transfer function identified, and the cost function can be rewritten as:

$$\hat{\mathbf{c}} = \min_{\mathbf{c} \in \mathbb{R}^m} \|\mathbf{Y}_N - \boldsymbol{\phi}_N \mathbf{c}\|_2^2 + \gamma \mathbf{c}^T \mathbf{P}^{-1} \mathbf{c} + \alpha \|\mathbf{c}\|_1. \quad (10)$$

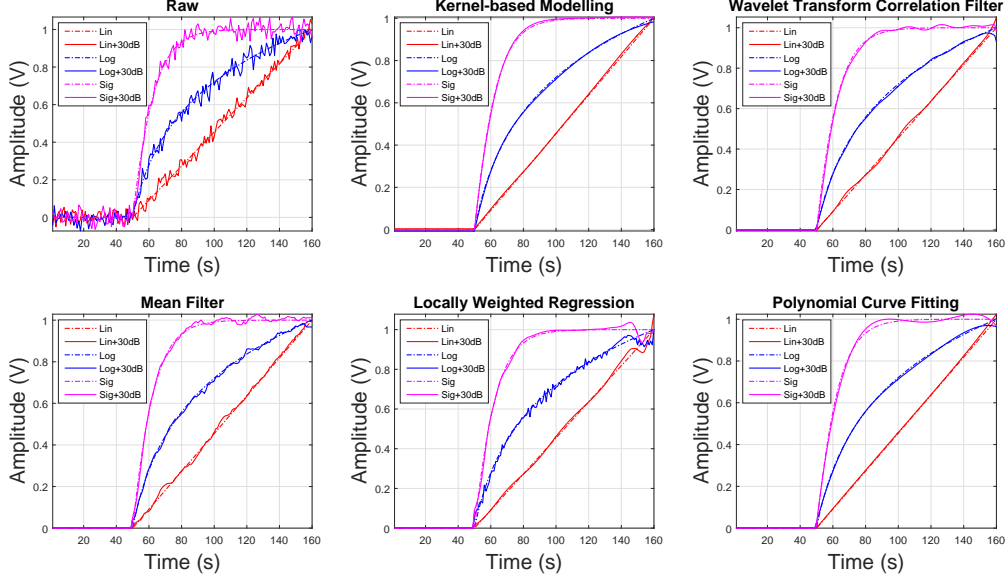
where  $\alpha$  is a positive coefficient to control the trade off between  $\mathcal{L}_1$  regulariser and kernel regulariser  $\gamma \mathbf{c}^T \mathbf{P}^{-1} \mathbf{c}$ .

The above equation can be considered as a special case of elastic net [42] which the  $\mathcal{L}_2$  norm regularisation is weighted by kernel matrix  $\mathbf{P}^{-1}$ . We here rearrange Eq.(10) and define two new parameters as:

$$\boldsymbol{\phi}_N^* = \frac{1}{\sqrt{1+\gamma}} \begin{bmatrix} \boldsymbol{\phi}_N \\ \sqrt{\gamma} \mathbf{B} \end{bmatrix}, \quad (11)$$

where  $\mathbf{B}$  is the upper triangular matrix from Cholesky factorisation of kernel matrix  $\mathbf{P}^{-1}$  ( $\mathbf{P}$  is symmetric) and

$$\mathbf{Y}_N^* = \begin{bmatrix} \mathbf{Y}_N \\ \mathbf{0} \end{bmatrix}. \quad (12)$$



**Figure 4:** Simulated Response Comparison for Different Data Pre-processing Methods.

Then, the cost function Eq.(10) can be written as:

$$\hat{c} = \min_{c^* \in \mathbb{R}^m} \|\mathbf{Y}_N^* - \phi_N^* c^*\|_2^2 + \alpha \|c^*\|_1, \quad (13)$$

where  $c^*$  is defined as:

$$c^* = \sqrt{1 + \gamma} c. \quad (14)$$

Due to the limitation of the input signal, the input matrix  $\phi_N^T \phi_N$  is not orthogonal.

As efficient algorithms for solving wide classes of convex optimization problems, interior-point methods are always efficient in terms of computation time and resource consumptions [43]. The Eq.(13) is convex but not differentiable. The  $\mathcal{L}_1$  regularisation LSP (Least Squares Problems) can be transformed to a convex quadratic problem, with linear inequality constraints. Therefore, we adopt an interior-point method (primal-dual interior-point method) [44] for this  $\mathcal{L}_1$  norm regularisation for  $\hat{c}^*$ . As the result,  $\hat{c}^*$  can be restored to  $\hat{c}$  as:

$$\hat{c} = \frac{1}{\sqrt{1 + \gamma}} \hat{c}^*. \quad (15)$$

## 4. Experimental Results and Discussion

### 4.1. Data Pre-processing Simulation Results

The data sets list in Table. 1 come from the average of 100,000 times simulations, which show the goodness-of-fit test results for three test signals. According to this table, although PCF achieves higher goodness-of-fit results for linear signal (SNR from 20dB to 50dB), the proposed non-parametric KBM method can also achieve quite close results. As for the logarithmic signal and sigmoid signal, the proposed method achieves higher goodness-of-fit results than the other methods, which indicates the signals processed by the proposed method are more close to the original signals. Therefore, the proposed non-parametric KBM method provides more reliable pre-processing results. Moreover, since the original test signals (linear signal,

logarithmic signal and sigmoid signal) are smooth and noiseless, when using these signals to evaluate the different data pre-processing methods, the method achieves higher goodness-of-fit is considered as a more smooth method. In addition, the simulated response results (see Fig. 4) for different data pre-processing methods indicate that the proposed method can provide more smooth results compared with other pre-processing methods.

### 4.2. Perfume Test Results

Based on the different data pre-processing methods, perfume I is used as an example to show the waveforms for the sensors which have response to the test samples, and results are plotted in Fig. 5. From this figure, it can be more directly seen that the proposed non-parametric KBM method can provide more smooth response waveforms compared with the other methods. Therefore, the proposed method can achieve much more reliable derivative-related features.

In odour classification stage, the features are often extracted from the data recorded until the steady-state response of gas sensors reached. Feature extraction methods generally fall into two categories: the human-supervised extraction based on expert knowledge and the automatic feature extraction methods that are completely data-driven. In this study, we have extracted the features by using a human-supervised method as the proof-of-concept of the presented method. Considering the derivative-based features are sensitive to noises, to demonstrate the effectiveness of our proposed method in dealing with noised signals, we in particular chose six derivative-based features commonly used by other works [45, 46]: the response of the maximum 1<sub>st</sub>-order derivative ( $D_{res}$ ), the response of the maximum 2<sub>nd</sub>-order derivative ( $D_{resx}$ ), the response of the minimum 2<sub>nd</sub>-order derivative ( $D_{resn}$ ), time interval between gas-in and maximum 1<sub>st</sub>-order derivative of response ( $t_{D_{res}}$ ), the time interval between gas-in and maximum 2<sub>nd</sub>-order derivative of response ( $t_{D_{resx}}$ ), and the time interval between gas-in and minimum 2<sub>nd</sub>-order derivative of response ( $t_{D_{resn}}$ ). The diagram of these features is shown in Fig. 6.



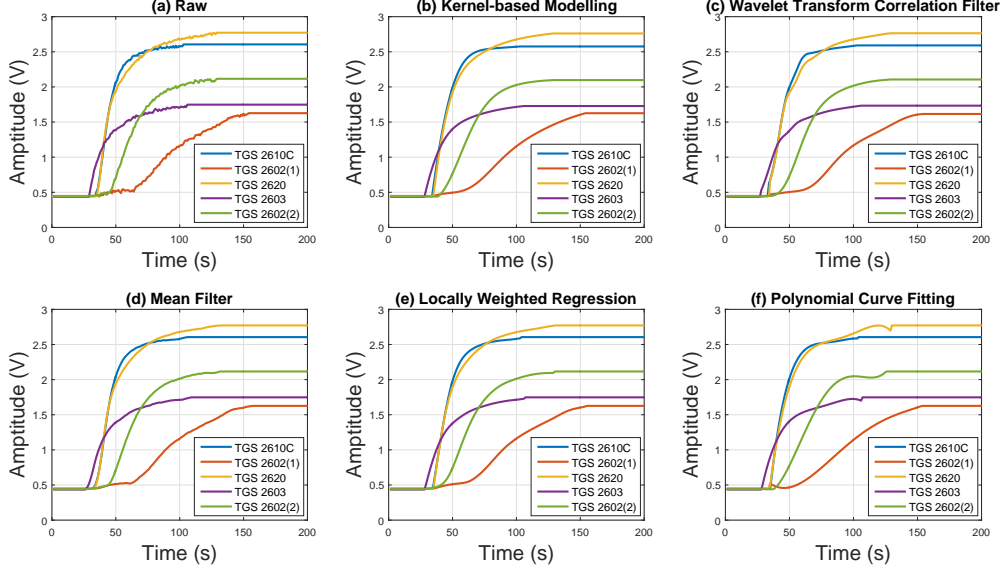


Figure 5: Data pre-processing Results for Perfume I.

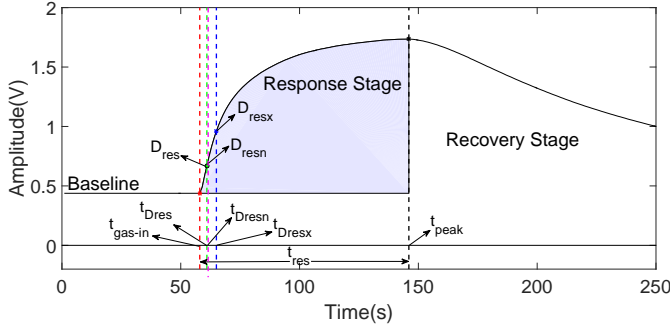


Figure 6: Features from the signal used for classification. Red line represents the time of gas-in,  $t_{gas-in}$ . The green line is the time point which has maximum 1<sup>st</sup>-order derivative of response,  $t_{Dres}$ . The pink line is the time point which has minimum 2<sup>nd</sup>-order derivative of response,  $t_{Dresn}$ . The blue line is the time point which has maximum 2<sup>nd</sup>-order derivative of response,  $t_{Dresx}$ . The black line is the time point which has maximum value of response,  $t_{peak}$ .

All the derivative-related features listed in Fig. 6 are extracted based on the six different pre-processed datasets. Then the Coefficient of Variation (CoV:  $Var_{coef}(x)$ ) assessment method was used to evaluate the stability of these features. According to Eq. 16, a smaller CoV value will indicate the better feature stability:

$$Var_{coef}(x) = \frac{\sigma(x)}{\bar{x}} \times 100\%, \quad (16)$$

where  $\sigma(x)$  is the standard deviation of the feature  $x$ , and  $\bar{x}$  is the mean value of the feature  $x$ .

Based on the CoV analysis results (listed in Table 2) for features of different types of perfumes, the proposed non-parametric KBM method has a smaller CoV value compared with the other methods, which means this method could provide much stable feature datasets in terms of improving the performance of the classification task.

#### 4.3. Classification Results

Based on different data pre-processing methods (non-parametric KBM, WPCF, MF, PCF and LWR), we extracted six derivative-related key features for the classification of the two types of perfumes by using the SVM classifier with RBF kernel. We used the five-fold cross-validation method for classification. The full dataset is partitioned into training and test sets by randomly selecting 20% of the data from each sample to form the test set and the remaining data to form the training set. This process was performed five times: where each data sample appears exactly once in a test set. The five-fold cross-validation method is run ten times and the averaged accuracy is used to assess the perfume classification accuracy.

The classification results are listed in Table 3. Without any pre-processing methods, the SVM classifier achieved a 62.57% average accuracy of classification. When the odour dataset was processed by WPCF, we achieved 94.12% average classification accuracy. MF pre-processing method for SVM classifier has a 68.57% average classification accuracy. PCF and LWR pre-processing methods attained 87.80% and 71.50% average classification accuracy, respectively. According to these classification results, the proposed non-parametric KBM method achieved a much higher average classification accuracy (96.23%) than other methods. The confusion matrix for the classification accuracy based on different data pre-processing methods is shown in Fig 7. The classification accuracy clearly indicates that the proposed data pre-processing method can provide more stable features to improve the performance of the odour classifier.

## 5. Conclusion

A novel non-parametric KBM odour data pre-processing method has been presented and its effectiveness has been tested on the recently developed NOS.E system. According to the test results, when extracting derivative-related features, the proposed non-parametric KBM method provides more reliable and stable pre-processing results comparing with the other pre-processing methods. Based on these derivative-related features,

**Table 1:** goodness-of-fit Result

Method	Test Signal	goodness-of-fit Result				
		Signal-to-Noise Ratio				
		10dB	20dB	30dB	40dB	50dB
N/A	Linear	0.0567	0.7016	0.9056	0.9702	0.9906
N/A	Logarithmic	0.1713	0.7379	0.9171	0.9738	0.9917
N/A	Sigmoid	0.2970	0.7777	0.9297	0.9778	0.9930
KBM	Linear	<b>0.8411</b>	0.9492	0.9818	0.9900	0.9912
KBM	Logarithmic	<b>0.8605</b>	<b>0.9553</b>	<b>0.9843</b>	<b>0.9916</b>	<b>0.9927</b>
KBM	Sigmoid	<b>0.8815</b>	<b>0.9620</b>	<b>0.9859</b>	<b>0.9914</b>	<b>0.9922</b>
WCTF	Linear	0.6740	0.8950	0.9605	0.9661	0.9635
WCTF	Logarithmic	0.7139	0.9063	0.9568	0.9670	0.9696
WCTF	Sigmoid	0.7557	0.9137	0.9573	0.9633	0.9638
MF	Linear	0.7057	0.9066	0.9694	0.9875	0.9912
MF	Logarithmic	0.7414	0.9179	0.9730	0.9887	0.9919
MF	Sigmoid	0.7802	0.9296	0.9750	0.9860	0.9876
LWR	Linear	0.7579	0.9256	0.9766	0.9923	0.9969
LWR	Logarithmic	0.7936	0.9349	0.9784	0.9905	0.9927
LWR	Sigmoid	0.8253	0.9432	0.9772	0.9841	0.9850
PCF	Linear	0.8409	<b>0.9498</b>	<b>0.9841</b>	<b>0.9950</b>	<b>0.9984</b>
PCF	Logarithmic	0.8598	0.9542	0.9815	0.9873	0.9881
PCF	Sigmoid	0.8783	0.9538	0.9711	0.9735	0.9738

**Note:** N/A: Not Applicable; KBM: Kernel-based modelling; WCTF: Wavelet Transform Correlation Filter; MF: Mean Filter; LWR: Locally Weighted Regression; PCF: Polynomial Curve Fitting.

NOS.E system can detect and identify two different perfumes with a 96.23% classification accuracy by using popular SVM classifier.

Further study will focus on the areas below: 1. Collecting more different samples to validate the implementation of the NOS.E system across various odour identification and classification applications; 2. Improving the efficiency and functionality of the NOS.E system by more research outputs with the proposed non-parametric KBM data pre-processing method and NOS.E hardware design optimisation; 3. Considering discriminative models for feature extraction [47]; 4. Applying some cost-sensitive classification algorithms to improve the performance of NOS.E system [48].

## Acknowledgement

This project was supported by the Wildlife Crime Tech Challenge (WCTC) and funded by the United States Agency for International Development (USAID). The authors would like to thank the Editor and the anonymous reviewers for their valuable comments and suggestions.

## References

- [1] A. D. Wilson, M. Baietto, Applications and advances in electronic-nose technologies, *Sensors* 9 (2009) 5099–5148.
- [2] J. W. Gardner, P. N. Bartlett, A brief history of electronic noses, *Sensors and Actuators B: Chemical* 18 (1994) 210–211.
- [3] A. Loutfi, S. Coradeschi, G. K. Mani, P. Shankar, J. B. B. Rayappan, Electronic noses for food quality: A review, *Journal of Food Engineering* 144 (2015) 103–111.
- [4] M. Peris, L. Escuder-Gilbert, A 21st century technique for food control: Electronic noses, *Analytica chimica acta* 638 (2009) 1–15.
- [5] I. Rodriguez-Lujan, J. Fonollosa, A. Vergara, M. Homer, R. Huerta, On the calibration of sensor arrays for pattern recognition using the minimal number of experiments, *Chemometrics and Intelligent Laboratory Systems* 130 (2014) 123–134.
- [6] E. Gobbi, M. Falasconi, G. Zambotti, V. Sberveglieri, A. Pulvirenti, G. Sberveglieri, Rapid diagnosis of enterobacteriaceae in vegetable soups by a metal oxide sensor based electronic nose, *Sensors and Actuators B: Chemical* 207 (2015) 1104–1113.
- [7] A. Caron, N. Redon, F. Thevenet, B. Hanoune, P. Coddeville, Performances and limitations of electronic gas sensors to investigate an indoor air quality event, *Building and Environment* 107 (2016) 19–28.

- [8] K. Brudzewski, S. Osowski, A. Dwulit, Recognition of coffee using differential electronic nose, *IEEE Transactions on Instrumentation and Measurement* 61 (2012) 1803–1810.
- [9] L. Zhang, F.-C. Tian, X.-W. Peng, X. Yin, A rapid discreteness correction scheme for reproducibility enhancement among a batch of mos gas sensors, *Sensors and Actuators A: Physical* 205 (2014) 170–176.
- [10] L. Zhang, D. Zhang, Domain adaptation extreme learning machines for drift compensation in e-nose systems, *IEEE Transactions on Instrumentation and Measurement* 64 (2015) 1790–1801.
- [11] J. L. Herrero, J. Lozano, J. P. Santos, J. I. Suárez, On-line classification of pollutants in water using wireless portable electronic noses, *Chemosphere* 152 (2016) 107–116.
- [12] H. Sun, F. Tian, Z. Liang, T. Sun, B. Yu, S. X. Yang, Q. He, L. Zhang, X. Liu, Sensor array optimization of electronic nose for detection of bacteria in wound infection, *IEEE Transactions on Industrial Electronics* (2017).
- [13] E. Westenbrink, R. P. Arasaradnam, N. O’Connell, C. Bailey, C. Nwokolo, K. D. Bardhan, J. Covington, Development and application of a new electronic nose instrument for the detection of colorectal cancer, *Biosensors and Bioelectronics* 67 (2015) 733–738.
- [14] J.-H. Kim, J. Chun, J. W. Kim, W. J. Choi, J. M. Baik, Self-powered, room-temperature electronic nose based on triboelectrification and heterogeneous catalytic reaction, *Advanced Functional Materials* 25 (2015) 7049–7055.
- [15] L. Gil-Sánchez, J. Soto, R. Martínez-Máñez, E. Garcia-Breijo, J. Ibáñez, E. Llobet, A novel humid electronic nose combined with an electronic tongue for assessing deterioration of wine, *Sensors and Actuators A: Physical* 171 (2011) 152–158.
- [16] D. Li, T. Lei, S. Zhang, X. Shao, C. Xie, A novel headspace integrated e-nose and its application in discrimination of chinese medical herbs, *Sensors and Actuators B: Chemical* 221 (2015) 556–563.
- [17] A. Somov, E. F. Karpov, E. Karpova, A. Suchkov, S. Mironov, A. Karelin, A. Baranov, D. Spirjakin, Compact low power wireless gas sensor node with thermo compensation for ubiquitous deployment, *IEEE Transactions on Industrial Informatics* 11 (2015) 1660–1670.
- [18] Z. Zhao, F. Tian, H. Liao, X. Yin, Y. Liu, B. Yu, A novel spectrum analysis technique for odor sensing in optical electronic nose, *Sensors and Actuators B: Chemical* 222 (2016) 769–779.
- [19] L. Zhang, Y. Liu, P. Deng, Odor recognition in multiple e-nose systems with cross-domain discriminative subspace learning, *IEEE Transactions on Instrumentation and Measurement* 66 (2017) 1679–1692.
- [20] X. Peng, L. Zhang, F. Tian, D. Zhang, A novel sensor feature extraction based on kernel entropy component analysis for discrimination of indoor air contaminants, *Sensors and Actuators A: Physical* 234 (2015) 143–149.
- [21] X. Sun, L. Liu, Z. Wang, J. Miao, Y. Wang, Z. Luo, G. Li, An optimized multi-classifiers ensemble learning for identification of ginsengs based on electronic nose, *Sensors and Actuators A: Physical* 266 (2017) 135–144.
- [22] S. Jiang, J. Wang, Y. Wang, S. Cheng, A novel framework for analyzing mos e-nose data based on voting theory: Application to evaluate the internal quality of chinese pecans, *Sensors and Actuators B: Chemical* 242 (2017) 511–521.
- [23] Y.-Q. Jing, Q.-H. Meng, P.-F. Qi, M.-L. Cao, M. Zeng, S.-G. Ma, A bioinspired neural network for data processing in an electronic nose, *IEEE Transactions on Instrumentation and Measurement* 65 (2016) 2369–2380.
- [24] X. Jiang, P. Jia, R. Luo, B. Deng, S. Duan, J. Yan, A novel electronic nose learning technique based on active learning: Eqbc-rbfn, *Sensors and Actuators B: Chemical* 249 (2017) 533–541.
- [25] H. Jamalabadi, A. Mani-Varnosfaderani, N. Alizadeh, Detection of alkyl

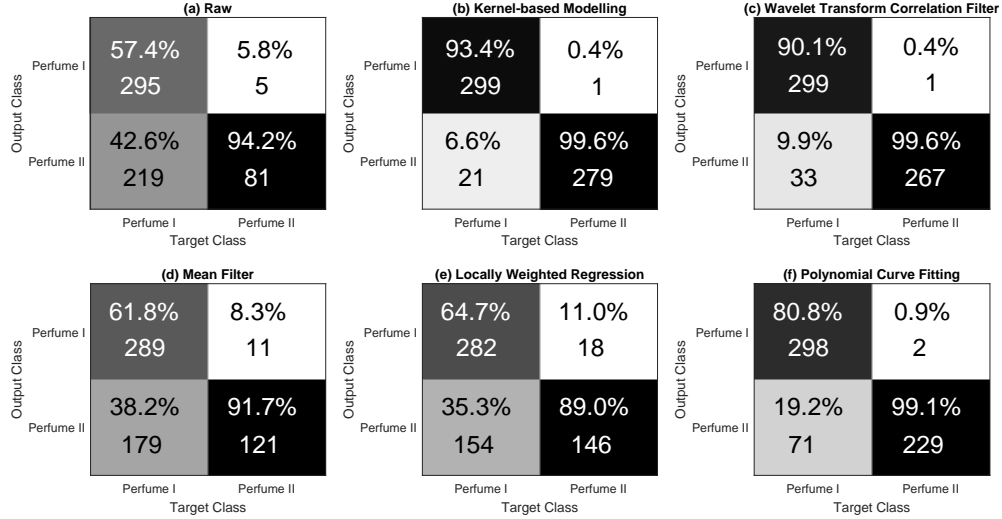
**Table 2:** The Odour Features' Coefficient of Variation for Perfume I and Perfume II

Features	Perfume I						Perfume II					
	KBM	WTCF	MF	LWR	PCF	N/A	KBM	WTCF	MF	LWR	PCF	N/A
TGS 2610C $D_{res}$	1.64	19.85	1.94	<b>1.22</b>	17.45	41.18	1.40	29.94	2.27	<b>1.13</b>	19.65	47.25
TGS 2610C $D_{resx}$	<b>0.71</b>	11.16	4.10	1.49	28.24	12.31	<b>0.72</b>	7.05	4.32	1.49	16.13	14.42
TGS 2610C $D_{resn}$	0.90	56.94	1.82	<b>0.84</b>	57.13	34.82	<b>0.61</b>	66.92	1.87	0.72	43.51	70.85
TGS 2610C $t_{Dres}$	<b>0.59</b>	6.32	2.28	0.80	33.66	31.17	<b>0.66</b>	3.42	2.59	0.97	33.87	31.48
TGS 2610C $t_{Dresx}$	<b>2.22</b>	53.52	12.31	14.02	26.18	48.91	<b>2.58</b>	82.42	14.96	15.59	30.55	50.81
TGS 2610C $t_{Dresn}$	11.07	18.50	17.10	57.13	29.37	<b>15.93</b>	18.69	21.88	18.14	56.91	<b>16.13</b>	17.70
TGS 2602I $D_{res}$	<b>1.91</b>	112.57	10.99	23.57	83.36	37.07	<b>2.44</b>	90.16	11.35	17.94	54.53	81.65
TGS 2602I $D_{resx}$	<b>1.68</b>	66.50	10.27	10.77	17.66	22.97	<b>1.85</b>	52.55	11.12	11.20	19.49	39.82
TGS 2602I $D_{resn}$	<b>1.58</b>	43.97	16.55	48.21	60.17	31.33	<b>1.79</b>	33.04	17.94	51.49	67.42	33.99
TGS 2602I $t_{Dres}$	<b>10.90</b>	30.30	17.52	120.75	37.24	16.510	<b>13.35</b>	33.49	19.31	95.21	15.38	19.25
TGS 2602I $t_{Dresx}$	<b>3.43</b>	103.43	14.50	34.42	109.80	43.46	<b>2.54</b>	73.34	13.09	20.40	64.10	80.94
TGS 2602I $t_{Dresn}$	<b>2.64</b>	85.17	16.38	8.09	18.11	22.66	3.67	87.15	17.98	<b>3.03</b>	15.61	41.02
TGS 2620 $D_{res}$	<b>4.00</b>	47.97	18.04	62.07	60.17	30.60	<b>4.92</b>	39.21	18.67	63.21	66.40	31.17
TGS 2620 $D_{resx}$	3.70	6.46	<b>3.50</b>	3.78	6.32	6.54	2.56	3.63	<b>2.39</b>	2.55	3.23	3.67
TGS 2620 $D_{resn}$	2.94	3.66	<b>2.79</b>	3.04	3.69	3.69	3.00	4.02	<b>2.85</b>	3.05	4.28	4.06
TGS 2620 $t_{Dres}$	3.94	4.36	<b>3.69</b>	3.89	4.33	4.41	4.60	5.17	<b>4.29</b>	4.56	5.12	5.23
TGS 2620 $t_{Dresx}$	3.34	4.18	<b>3.13</b>	3.52	5.15	4.23	2.93	4.00	<b>2.74</b>	3.15	4.49	4.04
TGS 2620 $t_{Dresn}$	1.12	6.78	2.05	<b>0.95</b>	5.63	6.04	1.19	6.57	2.25	<b>0.73</b>	4.60	9.89
TGS 2603 $D_{res}$	<b>0.52</b>	27.88	4.40	2.61	71.53	26.75	<b>0.40</b>	21.38	4.35	4.00	5.49	23.89
TGS 2603 $D_{resx}$	<b>0</b>	7.65	1.63	0.56	21.44	7.93	0.40	5.11	1.54	<b>0.22</b>	8.90	10.12
TGS 2603 $D_{resn}$	<b>0</b>	8.21	2.53	1.49	9.30	15.89	<b>0</b>	4.48	2.69	1.64	6.92	12.92
TGS 2603 $t_{Dres}$	<b>0</b>	6.76	3.66	0.76	5.93	43.86	<b>0</b>	5.35	3.98	1.02	5.24	40.47
TGS 2603 $t_{Dresx}$	<b>4.66</b>	5.65	36.88	62.61	69.01	44.10	8.71	<b>5.49</b>	37.96	65.71	5.49	42.65
TGS 2603 $t_{Dresn}$	<b>3.03</b>	36.98	5.91	58.37	79.67	53.22	<b>0</b>	14.10	3.16	37.05	22.34	46.67
TGS 2602II $D_{res}$	<b>0</b>	10.71	9.56	0.53	36.24	49.88	<b>0</b>	8.50	4.41	0.70	38.62	53.05
TGS 2602II $D_{resx}$	<b>0.10</b>	12.20	7.82	52.17	15.26	41.60	<b>0.06</b>	5.53	8.93	55.90	14.76	39.37
TGS 2602II $D_{resn}$	<b>0.66</b>	20.53	12.96	9.22	60.15	38.52	<b>0.43</b>	20.77	9.78	16.62	53.35	39.57
TGS 2602II $t_{Dres}$	<b>1.33</b>	43.22	17.30	42.88	65.11	48.52	<b>3.46</b>	17.41	13.45	44.96	34.93	40.79
TGS 2602II $t_{Dresx}$	<b>2.52</b>	17.24	33.01	5.96	9.05	40.38	3.82	18.06	26.64	<b>2.11</b>	6.67	44.34
TGS 2602II $t_{Dresn}$	<b>0</b>	9.60	23.55	38.84	76.17	39.82	<b>0</b>	5.75	26.79	40.32	75.87	40.30

**Table 3:** Classification Results

Pre-processing Method	Parameters of Pre-processing Method	Average Accuracy of SVM Classification (%)
N/A	N/A	62.57
Kernel-based modelling method	Kernel = RBF, $\rho = 0.0005$ , $m = 200$	96.23
Wavelet Package Correlation Filter	Wavelet = Daubechies - 3, Level = 5	94.12
Mean Filter	Window width = 7	68.57
Polynomial Curve Fitting	Order = 3	87.80
Locally Weighted Regression	$\tau = 5$	71.50

- amine vapors using ppy-zno hybrid nanocomposite sensor array and artificial neural network, *Sensors and Actuators A: Physical* 280 (2018) 228–237.
- [26] S. Cui, J. Wu, J. Wang, X. Wang, Discrimination of american ginseng and asian ginseng using electronic nose and gas chromatography–mass spectrometry coupled with chemometrics, *Journal of ginseng research* 41 (2017) 85–95.
- [27] J. de Swart, N. van Gaal, D. J. Berkhout, T. G. de Meij, N. K. de Boer, Smoking influences fecal volatile organic compounds composition, *Clinical Gastroenterology and Hepatology* 16 (2018) 1168–1169.
- [28] S. Dragonieri, G. Pennazza, P. Carratu, O. Resta, Electronic nose technology in respiratory diseases, *Lung* 195 (2017) 157–165.
- [29] W. Yang, J. Yu, F. Pei, A. M. Mariga, N. Ma, Y. Fang, Q. Hu, Effect of hot air drying on volatile compounds of flammulina velutipes detected by hs-spmc–gc–ms and electronic nose, *Food chemistry* 196 (2016) 860–866.
- [30] S. Grassi, E. Casiraghi, S. Benedetti, C. Alamprese, Effect of low-protein diets in heavy pigs on dry-cured ham quality characteristics, *Meat science* 131 (2017) 152–157.
- [31] L. Xu, X. Yu, L. Liu, R. Zhang, A novel method for qualitative analysis of edible oil oxidation using an electronic nose, *Food chemistry* 202 (2016) 229–235.
- [32] F. Röck, N. Barsan, U. Weimar, Electronic nose: current status and future trends, *Chemical reviews* 108 (2008) 705–725.
- [33] E. Hines, E. Llobet, J. Gardner, Electronic noses: a review of signal processing techniques, *IEE Proceedings-Circuits, Devices and Systems* 146 (1999) 297–310.
- [34] Z. Sun, C. Chang, Structural damage assessment based on wavelet packet transform, *Journal of structural engineering* 128 (2002) 1354–1361.
- [35] G. Korotcenkov, B. Cho, Instability of metal oxide-based conductometric gas sensors and approaches to stability improvement (short survey), *Sensors and Actuators B: Chemical* 156 (2011) 527–538.
- [36] S. K. Jha, R. Yadava, Denoising by singular value decomposition and its application to electronic nose data processing, *IEEE Sensors Journal* 11 (2011) 35–44.
- [37] Z. Liang, F. Tian, C. Zhang, H. Sun, A. Song, T. Liu, Improving the robustness of prediction model by transfer learning for interference suppression of electronic nose, *IEEE Sensors Journal* (2017).
- [38] G. Pillonetto, G. De Nicolao, A new kernel-based approach for linear system identification, *Automatica* 46 (2010) 81–93.
- [39] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, L. Ljung, Kernel methods in system identification, machine learning and function estimation: A survey, *Automatica* 50 (2014) 657–682.
- [40] M. M. Bordbar, J. Tashkhourian, B. Hemmateenejad, Qualitative and quantitative analysis of toxic materials in adulterated fruit pickle samples by a colorimetric sensor array, *Sensors and Actuators B: Chemical* 257 (2018) 783–791.
- [41] L. Ye, A. Argha, H. Yu, B. G. Celler, H. T. Nguyen, S. Su, Dynamic characteristics of oxygen consumption, *Biomedical engineering online* 17 (2018) 44.
- [42] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2005) 301–320.
- [43] H. Lebrecht, S. Boyd, Antenna array pattern synthesis via convex optimization, *IEEE transactions on signal processing* 45 (1997) 526–532.
- [44] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, An interior-point method for large-scale  $\ell_1$  regularized least squares, *IEEE journal of selected topics in signal processing* 1 (2007) 606–617.
- [45] T. Eklöv, P. Mårtensson, I. Lundström, Selection of variables for interpreting multivariate gas sensor data, *Analytica Chimica Acta* 381 (1999) 221–232.
- [46] M. Pardo, G. Sberveglieri, Comparing the performance of different features in sensor arrays, *Sensors and Actuators B: Chemical* 123 (2007) 437–443.
- [47] L. Zhang, X. Wang, G.-B. Huang, T. Liu, X. Tan, Taste recognition in e-tongue using local discriminant preservation projection, *IEEE Transactions on Cybernetics* (2018).



**Figure 7:** The confusion matrix for the classification accuracy for different data-preprocessing methods.

- [48] L. Zhang, D. Zhang, Evolutionary cost-sensitive extreme learning machine, IEEE transactions on neural networks and learning systems 28 (2017) 3045–3060.

## Biographies

**Wentian Zhang** received the B.E. degree from Nanjing Agriculture University (NJAU), China, in 2010, and the M.E. degree from Shanghai Maritime University, China, in 2012. He is currently pursuing the Ph.D. degree in Faculty of Engineering and IT from University of Technology Sydney (UTS), Australia. His current research interests include electronic nose system, Machine Learning, signal processing, and medical devices.

**Taoping Liu** received the B.E. degree from Central South University (CSU), China, in 2013, and the M.E. degree from Xi'an Jiaotong University (XJTU), China, in 2016. He is currently pursuing the Ph.D. degree in Faculty of Engineering and IT from University of Technology Sydney (UTS), Australia. His current research interests include Machine Learning, condition monitoring, and signal processing.

**Lin Ye** was born in Liaoning, China, in 1989. He received the B.E. (Hons. I) degree in electrical engineering from the University of Technology Sydney (UTS), Sydney, Australia, in 2013, where he is currently pursuing the Ph.D. degree with the Center for Health Technologies. His current research interests include system identification, modelling of cardiorespiratory response, and calibration of inertial measurement unit.

**Maiken Ueland** received her B.Sc. with Honours in Forensic Science, with a double minor in 2012 from the University of Ontario Institute of Technology, Canada. She then moved on to do a PhD at University of Technology Sydney (UTS), Australia, which was completed in 2016. She currently works as a Chancellor's Postdoctoral Research Fellow at UTS where she is interested in enhancing time since death estimations of human remains and the development of electronic nose systems.

**Shari L. Forbes** is a Professor and former ARC Future Fellow in the Centre for Forensic Science at the University of Technology Sydney (UTS), Australia. Prof. Forbes completed a B.Sc. with Honours in Applied Chemistry and Forensic Science and a PhD (Forensic Chemistry) at UTS. She is a Fellow of the Royal Society of NSW, Secretary-General of the Australian Academy of Forensic Sciences, Vice-President of the NSW Branch of the Australian and New Zealand Forensic Science Society (ANZFSS), and the Pacific Officer for the Initiative on Forensic Geology, a directive of the International Union of Geological Sciences. Her research investigates the chemical odour profile of forensic evidence as it relates to detector dogs in law enforcement.

**Steven W. Su** received both the B.S. and M.S. degrees from Harbin Institute of Technology (HIT), China, in 1990 and 1993, respectively, and the Ph.D. degree from the Research School of Information Sciences and Engineering, Australian National University (ANU), Australia, in 2002. He is currently an Associate Professor in the Faculty of Engineering and IT, University of Technology Sydney (UTS), Australia. His main research interests include biomedical system modelling and control, rehabilitation engineering, navigation system design, robust and adaptive control, fault tolerant control, and wearable monitoring system.